

E164 Homework 1

January 19, 2022

1 Computational Biology and Python

Resources:

1. Computer Setup: If you are having issues installing or running any code, please review the setup [document](#) again. If you followed all steps correctly but still cannot run the jupyter notebooks, post on Piazza.
2. w1_python_intro.ipynb: If you are new to Python, you will find this notebook a good starting point. The notebook also contains code that you may want to re-use for the homework.
3. Python tutorials online: You are free to use any tutorials from the Internet on Numpy, Scipy, Pandas, or any other Python package that you may use to solve the homework problems.
4. Submit your homework to GradeScope by downloading the jupyter notebook as a PDF. Go to Files -> Download as -> PDF.

Due date: 25th January on GradeScope.

1.1 Problem 1: Computational Biology and E164

Summarize the paper Computational Systems Biology by Hiroaki Kitano ([PDF here](#)) in your own words. As you write the summary, make a note of terms/phrases/concepts that you did not fully understand from the paper. Also, reflect on things from the paper that intrigued you the most and things that you would like to discuss in E164.

Answer:

1.2 Problem 2: Practice your Python: Preliminary modeling of COVID-19 growth

- (a) Use the code in this notebook to download confirmed COVID-19 case data from Nov 1 2021 to now for California and New York. Plot this data on two different graphs with labeled axes (date on X-axis and case numbers on Y-axis) and legends.
- (b) Explore how the people of California and New York got their vaccines against COVID-19. Generate a plot for the number of people fully vaccinated against COVID-19 in these two states. Hint: You can use the **Pandas** package documentation to learn about the various functions available to use on a Pandas dataframe (that is, the dataframe `x` in the code given).
- (c) Model the infection growth by fitting a curve to the data obtained in part (a). You may try to fit an exponential function to the data or modify the function as needed to get the best fit possible. Report the mathematical function that you used to fit the data for the two states,

the optimized parameters, and their covariances. Comment on the optimality of your curve fit.

- (d) Compute the doubling time of the infection spread from your mathematical model in part (c).
- (e) Predict the date on which the number of confirmed cases in the two states (CA, NY) and the country (USA) will be equal to the number of confirmed cases on Nov 1 2021. You may assume that the end point of the available data is the peak of the infection growth and that the number of infections will go down in a mirrored fashion. Show your work on how you compute your prediction using Python. Note that you need to make three predictions, one each for CA, NY, and USA.
- (f) Plot your results in part (e) by making three different graphs which show the existing data and your forecast clearly. If this was the curve that you wanted to fit (Nov 1 - now - future), which function would you choose? Are there any other ways in Scipy to fit such a data? You don't need to actually fit this data but you are welcome to try for bonus credit.

1.2.1 Get real COVID-19 infection growth data using the following code:

To get the real COVID-19 infection data, you need to get the COVID-19 datahub package. Run the command `pip install covid19dh` to install this package from your terminal or search for the package `covid19dh` in your Anaconda Navigator. If you don't find it in the navigator, you can simply use the command in the next cell to install it. You may also need to install the `Pandas` package to run the code and work with the dataset. Simply run `pip install pandas` or search for `Pandas` in your navigator.

```
[1]: # Uncomment the following line to install covid19dh package:
# !pip install covid19dh
```

```
[1]: from covid19dh import covid19
from datetime import date
import numpy as np
# Use the covid19dh package to get the data
x, src = covid19("USA", level = 2, start = date(2021,11,1), verbose = False)
# Then, with some magic of the Pandas package we get the data out in numpy
↪arrays:
california_case_data = np.array(x[x['key_apple_mobility'] ==
↪'California']['confirmed'])
newyork_case_data = np.array(x[x['key_apple_mobility'] == 'New
↪York']['confirmed'])
```

```
[2]: # Solutions:
# (a)
```

```
[3]: # (b)
```

```
[4]: # (c)
```

[5]: # (d)

[6]: # (e)

[7]: # (f)

1.3 Problem 3: Feedback! The key to all things that function robustly.

After finishing the rest of the problem set, please write here the number of hours you needed to finish it. Also comment on things in the homework and the class lectures that you did not find interesting/took too long/difficult or anything else that you would see changed. Remember you can always use private anonymous posts on Piazza to send any feedback as well.

Answer: